IMPACT ASSESSMENT METHODOLOGIES FOR MICROFINANCE: A REVIEW

Professor David Hulme
Institute for Development Policy and Management
University of Manchester

[Paper prepared in conjunction with the AIMS Project for the Virtual Meeting of the CGAP Working Group on Impact Assessment Methodologies (April 17-19, 1997).]

May 1997

The meeting was organized by the AIMS Project. AIMS is funded by the Microenterprise Impact Project (PCE-0406-C-00-5036-00) of USAID's Office of Microenterprise Development. The Project is conducted through a contract with Management Systems International, in cooperation with the Harvard Institute for International Development, the University of Missouri, and The Small Enterprise Education and Promotion Network.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Tables

Figures

EXECUTIVE SUMMARY

1 This paper has been prepared for CGAP's Working Group on Impact Assessment (IA) Methodologies. It argues that improving IA studies will require both better 'science' and better 'art' (Section 1).

2 Impact assessments can be viewed as laying on a continuum between 'proving impacts' and 'improving practice'. While many different groups have interests in IA there is currently great pressure on donor agencies to prove impact (Section 2).

3 Different conceptual frameworks provide the foundations for all IAs. These must cover a model of the impact chain, the identification of units (levels) of assessment and the specification of the types of impact to be assessed (Section 3).

4 The range of methods used for IA have extended over the last decade - sample surveys, rapid appraisal, participant observation, case study and participatory learning and action (Table 2). Each method has particular strengths and weaknesses (Table 3). The choice of methods must be based on objectives, costs and feasibility. Mixed methods have much to recommend them (Section 4).

5 Attribution and fungibility are key problems for impact assessment studies. The ways in which scientific, humanities-based and participatory methods deal with these problems are examined. A variety of techniques to avoid or reduce these problems are reviewed (see summary in Appendix 6). It is argued that the problem of fungibility is not as great as many writers suggest. (Section 5)

6 Impact assessors must also design their studies so as to avoid a number of practical problems. These include the trade-offs between IA costs and rigor; respondent motivation; consultant and interviewer availability; and how to avoid low-impact IAs (Section 6).

7 The key to selecting an appropriate methodology rests on achieving a fit between IA objectives, program contexts and the availability of resources. Four types of methodology are identified - monitoring, simple, moderate and complex - and their key characteristics are examined (Section 7).

8 The conclusion presents a number of recommendations about how to improve IA practice (Section 8). These are:

- More emphasis should be placed on internal impact monitoring within MFIs, rather than on impact assessment;
- The idea of methodological rigour must be applied to all designs and methods;
- The main need in IA is to improve the quality of simple IA designs;
- Monitoring MFI performance is not sufficient - IAs must look at impacts on livelihoods;

- The Household Economic Portfolio Model (HEPM) and other multi-level approaches, are to be recommended for complex IA studies;

- Problems of attribution can be overcome by careful design. The problem of fungibility has been exaggerated;

- Many IAs have 'low impact' because they do not address the issue of dissemination sufficiently;

- Donor activity has done little to strengthen the human resources or institutions available for IA in non-OECD countries, this must be corrected; and

- Improving IA standards requires major efforts to raise the capacities of an 'army' of impact assessors; a code of practice, careful training and networking are needed.

IMPACT ASSESSMENT METHODOLOGIES FOR MICROFINANCE: A REVIEW

1. INTRODUCTION

This paper has been prepared for the Consultative Group to Assist the Poorest (CGAP) Working Group on Impact Assessment Methodologies. It reviews the methodological options for assessing the impacts of microfinance programs in the light of their costs, feasibility and IA objectives (see Appendix 1 for Terms of Reference). The recent rise of microfinance programs as a strategy for poverty reduction, enterprise development and economic growth makes this an important issue as the claims about the benefits generated by microfinance have outstripped the evidence that is presently available.

This review builds upon a number of commissioned background papers (Appendix 2), the discussions of a virtual meeting convened by CGAP (see Gaile 1997 and Appendix 3), the work of the Assessing the Impact of Microenterprise Services (AIMS) Project (Appendix 4), the wider literature (Appendix 5) and the author's personal research and experience. The paper views impact assessment (IA) as being '...as much an art as a science...' (a phrase lifted from Little 1997:2). Enhancing the contribution that impact assessment can make to developmental goals requires both better science and better art. The scientific improvements relate to improving standards of measurement, sampling and analytical technique. Econometricians and statisticians are particularly concerned with this field. In the future it might also involve the establishment of minimum standards (or a code of practice) for certain types of impact assessments. Improving the 'art' of impact assessment has at least three strands. One concerns making more systematic and informed judgements about the overall design of Ias in relation to their costs, specific objectives and contexts. The second is about what mixes of impact assessment methods are most appropriate for any given study. The third relates to increasing our understanding of the ways in which the results of IA studies influence MFI policy makers and managers.

The reader should be aware that this paper seeks to identify key issues and explore them to promote debate on how to improve IA: it does not claim to solve the inherent problems of IA!

2. IMPACT ASSESSMENT: OBJECTIVES

Impact assessment studies have recently become popular with donor agencies and, in consequence, have become an increasingly significant activity for recipient agencies. In part this reflects a cosmetic change, with the term IA simply being substituted for evaluation. But it has also been associated with a greater focus on the outcomes of interventions, rather than inputs and outputs. While the goals of IA studies commonly incorporate both 'proving' impacts and 'improving' interventions, IAs are more likely to prioritise the proving goal than did the evaluations of the 1980s. A set of factors are associated with the extreme 'pole' positions of this continuum and these underpin many of the issues that must be resolved (and personal and institutional tensions that arise) when impact assessments are being initiated (Figure 1).

Behind the shift from 'evaluation' to 'IA' are a number of factors. These are not explored in any detail in this paper but they form an essential element for the understanding of IA and its potential contributions. Explicitly, IAs are promoted by both the sponsors and implementers of programs so that they can learn

what is being achieved and improve the effectiveness and efficiency of their activities. Implicitly, IAs are a method by which sponsors seek to get more information about program effectiveness than is available from the routine accountability systems of implementing organisations. IAs are also of significance to aid agencies in terms of meeting the ever increasing accountability demands of their governments (in this era of 'results' and 'value for money') and for contesting the rhetoric of the anti-aid lobby. While recipient agencies benefit from this, they are one stage removed, and many are likely to see donor-initiated IA as an activity that has limited practical relevance for program activities. To quote the director of a large Asian microfinance institution that has received substantial amounts of aid financed IA consultancy and internal IA-capacity building '...impact assessment studies keep donors happy... we don't use them very much'.

Figure 1: The Goals of Impact Assessment

	PROVING<----------->IMPROVING	
	IMPACTS	PRACTICE
Primary Goal	Measuring as accurately as possible the impacts of an intervention	Understanding the processes of intervention and their impacts so as to improve those processes
Main Audiences	Academics and researchers Policymakers Evaluation departments Program Managers	Program managers Donor field staff NGO personnel Intended beneficiaries
Associated Factors	Objectivity Theory External Top down Generalisation Academic research Long timescales Degree of confidence	Subjectivity Practice Internal Bottom up Contextualisation Market research Short timescales Level of plausibility

A final issue to raise in this section is whether the expectations of OECD agencies about the feasibility of the accurate measurement of impacts in the difficult contexts of developing countries (limited numbers of professional researchers, few written records, illiteracy, communication problems etc) are higher than in their own countries. Looking out of my office window in central Manchester at an EU-financed 'small enterprise development' program I see no evidence of a serious approach to IA. If recipients perceive that the IA standards expected of 'them' are higher than donors expect of themselves then IA will be seen as an external imposition rather than a shared opportunity.

3. ASSESSING IMPACT: THE CHOICE OF CONCEPTUAL FRAMEWORKS

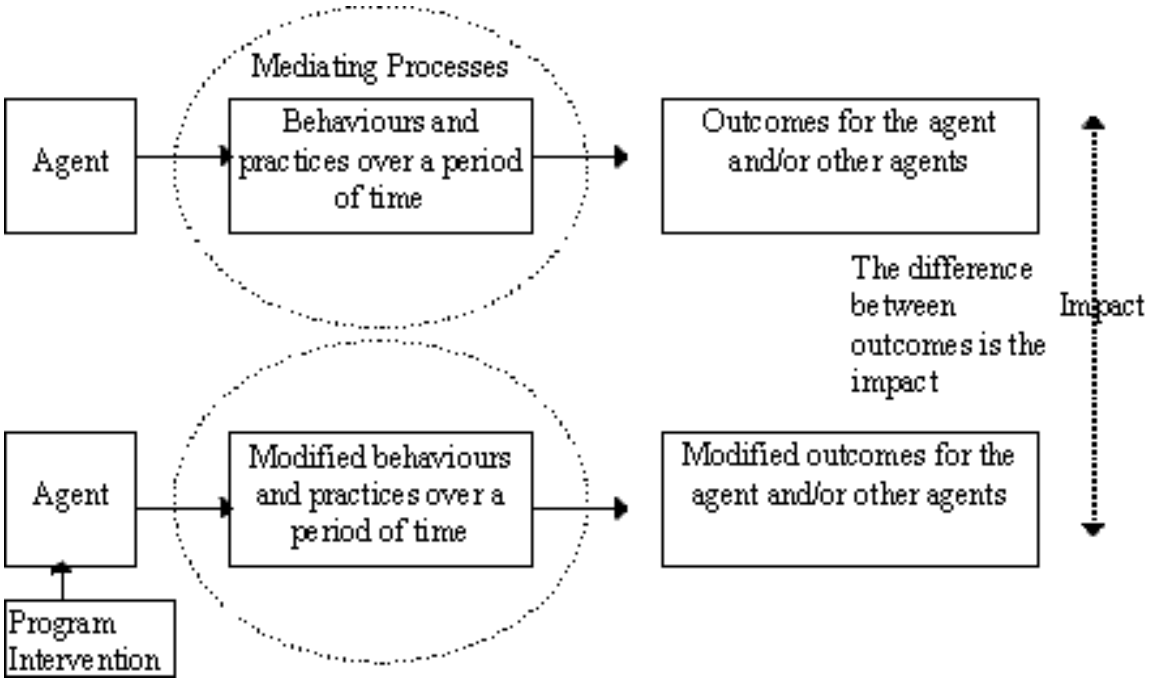
All impact assessment exercises have a conceptual framework at their heart. In well-resourced IAs with long 'lead-in' times such frameworks are usually explicitly identified (e.g.;Sebstad et al 1995; Schuler and Hashemi 1994). By contrast, in many smaller scale exercises the framework is implicit and may be seen as 'common sense'. There are three main elements to a conceptual framework:

(i) a model of the impact chain that the study is to examine;
(ii) the specification of the unit(s), or levels, at which impacts are assessed; and
(iii) the specification of the types of impact that are to be assessed.

3.1 Models of Impact Chains

Behind all microfinance programs is the assumption that intervention will change human behaviours and practices in ways that lead to the achievement (or raise the probability of achievement) of desired outcomes. IAs assess the difference in the values of key variables between the outcomes on 'agents' (individuals, enterprises, households, populations, policymakers etc) which have experienced an intervention against the values of those variables that would have occurred had there been no intervention (Figure 2). The fact that no agent can both experience an intervention and at the same time not experience an intervention generates many methodological problems. All changes are influenced by mediating processes (specific characteristics of the agent and of the economic, physical, social and political environment) that influence both behavioural changes and the outcomes in ways that are difficult to predict (Sebstad et al 1995).

Figure 2: The Conventional Model of the Impact Chain



The impact chain is very simply depicted above in Figure 2. A more detailed conceptualisation would present a complex set of links as each 'effect' becomes a 'cause' in its own right generating further effects. For example, in a conventional microfinance project a package of technical assistance and capital changes the behaviour (and products) of a microfinance institution (MFI). The MFI subsequently provides different services to a client, most commonly in the form of a loan. These services lead to the client modifying her/his microenterprise activities which in turn leads to increased/decreased microenterprise income. The change in microenterprise income causes changes in household income which in turn leads to greater/lesser household economic security. The modified level of household economic security leads to changes in the morbidity and mortality of household members, in educational and skill levels and in future economic and social opportunities. Ultimately, perhaps, these changes lead to modifications in social and political relations and structures. The complexity of such chains provides the assessor with a range of choices about which link (or links) to focus on. For microfinance, it is useful to distinguish between two main schools of thought with regard to which link(s) in the chain to focus on. For convenience, these are termed the 'intended beneficiary'[1] school and the 'intermediary' school.

[1] I use the term intended beneficiary here, rather than client, as most MFIs utilize (or have utilised) aid funds that are intended, at least in part, to benefit poor or vulnerable people and not purely self-selected clients. This is an important point as, (i) some MFIs present their client populations as intended beneficiaries when many clients are known to be non-poor, and (ii) some consultative groups present the client populations of MFIs as 'the poorest' when they know full well that only a proportion of clients are poor and that very few are 'the poorest' ie. social outcasts, destitute, disabled, refugees, widows or elderly.

The intended beneficiary school, building on the ideas of conventional evaluation, seeks to get as far down the impact chain as is feasible (in terms of budgets and techniques) and to assess the impact on intended beneficiaries (individuals or households). The intermediary school focuses purely on the beginning of the chain and in particular on changes in the MFI and its operations. Its roots are closely associated with the Ohio State School's analyses of rural finance. Generally, two key variables are focused on: institutional outreach and institutional sustainability (Yaron, Benjamin and Piprek 1997). If both outreach and sustainability have been enhanced then the intervention is judged to have a beneficial impact as it has widened the financial market in a sustainable fashion. This is based on the assumption that the institutional impacts will extend the choices of people looking for credit and savings services and that this extension of choice ultimately leads to improved microenterprise performance and household economic security. While this assumption can be supported by theoretical frameworks (if a set of further assumptions are made about perfect competition and other factors) it is an assumption which has proved invalid in a number of experiences.[2] In addition, it will not reveal borrower 'cross-financing' of loans (Wiig 1997) which may threaten the long term viability of an MFI.

While the choice between these two schools can ultimately be seen as an ideological choice (does one prioritise contributions to improved welfare or to more efficient markets?) it is possible to recognise different strengths and weaknesses. The intended beneficiary school makes fewer assumptions about the impact chain and is better able to distinguish 'who' benefits and 'how'. It is, however, demanding in both methodological and cost terms. The intermediary school usefully incorporates notions of sustainability and provides an IA methodological framework that can be operated largely with pre-existing data. It is, though, very weak on 'who' benefits and 'how' (as illustrated by assessments of the USAID-financed APPLE program[3]). Possible ways of strengthening the intermediary school approach have been suggested by Feinstein (1997) through the analysis of borrower transaction costs. He proposes the collection of longitudinal data on borrowers transaction costs (ibid:5) to assess whether an MFI has benefited borrowers. This offers a potential 'bridge' between the two main 'schools', if data on 'who' borrowers are also collected.

3.2 Units of Assessment

Following on from the design of a model of the impact path comes the choice of the unit(s) of assessment (or levels of assessment). Common units of assessment are the household, the enterprise or the institutional environment within which agents operate. Occasionally studies have attempted to assess impact at an individual level (e.g.;Goetz and Sen Gupta 1995; Peace and Hulme 1994), but this is relatively rare. More recently some studies have attempted to assess impacts at a number of levels, such as Hulme and Mosley

[2] The grandest invalidation is probably the UK in the mid to late 1980s when an unprecedented increase in outreach and profitability of institutions - providing financial services for households and small enterprises created a 'bubble' which ultimately destroyed many enterprises, led to some losing their homes and impoverished hundreds of thousands of households. Also note that Mayoux (1997) has found that programs performing well in terms of outreach and repayment rates can have negative impacts in terms of women's empowerment.

[3] I cannot locate my copy of a GEMINI study on the USAID Anti-Poverty Lending Program! This study ably computed almost every measure of outreach, outreach growth and institutional financial health possible: it failed to contain any information on 'who' clients were, but simply assumed they were 'the poor'.

(1996) who looked at microenterprise, household community and institutional levels. A comprehensive attempt at this is being made by USAID's AIMS Project. Through a household economic portfolio model (HEPM) it seeks to assess impacts at household, enterprise, individual and community levels and thus produce a fuller picture of overall impacts (Chen and Dunn 1996).

The relative advantages and disadvantages of different units of assessment are summarised in Table 1. As can be seen, a focus purely on the 'individual' or the 'enterprise' has such drawbacks that they could be viewed as discredited. The household economic portfolio model has much to recommend it - especially if institutional impacts are incorporated in the community level analysis. It does have the profound disadvantage, though, of making assessment demanding in terms of costs, skilled personnel and time. If used with limited resources it risks sacrificing depth for breadth of coverage of possible impacts.

Table 1: Units of Assessment and their Advantages and Disadvantages

Unit	Advantages	Disadvantages
Individual	# Easily defined and identified	# Most interventions have impacts beyond the individual # Difficulties of disaggregating group impacts and impacts on 'relations'
Enterprise	# Availability of analytical tools (profitability, return on investment etc)	# Definition and identification is difficult in microenterprises # Much microfinance is used for other enterprises and/or consumption # Links between enterprise performance and livelihoods need careful validation
Household	# Relatively easily defined and identified # Permits an appreciation of livelihood impacts # Permits an appreciation of interlinkages of different enterprises and consumption	# Sometimes exact membership difficult to gauge
Community	# Permits major externalities of interventions to be captured	# Quantitative data is difficult to gather # Definition of its boundary is arbitrary
Institutional Impacts	# Availability of data # Availability of analytical tools (profitability, SDIs, transaction costs)	# How valid are inferences about the outcomes produced by institutional activity?
Household Economic Portfolio (ie; household, enterprise, individual and community)	# Comprehensive coverage of impacts # Appreciation of linkages between different units	# Complexity # High costs # Demands sophisticated analytical skills # Time consuming

3.3 Types of Impact

An almost infinite array of variables can be identified to assess impacts on different units. To be of use these must be able to be defined with precision and must be measurable. Conventionally, economic indicators have dominated microfinance IAs with assessors particularly keen to measure changes in income despite the enormous problems this presents. Other popular variables have been levels and patterns of expenditure, consumption and assets. A strong case can be made that assets are a particularly useful indicator of impact because their level does not fluctuate as highly as other economic indicators and is not simply based on an annual estimate (Barnes 1996:v).

The social indicators that became popular in the early 1980s (e.g.; educational status, access to health services, nutritional levels, anthropometric measures and contraceptive use) have recently been extended into the socio-political arena in an attempt to assess whether microfinance can promote empowerment (Mayoux 1997; Goetz and Sen Gupta 1996; Schuler and Hashemi 1994; Hashemi et al 1995). This has led to the measurement of individual control over resources, involvement in household and community decision-making, levels of participation in community activities and social networks and electoral participation. The bulk of this work has focused on gender relations, but there are sometimes partially-formulated assessments of class relations within it (Fuglesang and Chandler 1993). These extensions to the types of impact assessed permit IAs to be more sophisticated and to shed light on developmental impacts at a time when the goals of development have also been extended. They do add, however, to the complexity of IA work and require the skills of assessors who are experienced at making judgements on social relations.

Sebstad et al (1995) usefully distinguish between 'domains of change' (e.g.; household income) and the specific 'markers of change' (e.g.; amount of income, number of income sources and seasonality of income) within each domain. While not fully comprehensive, the detailed sets of domains and markers, produced in their paper provide an excellent checklist for impact assessors to consider at the IA design stage.[4] Often the exact markers used will be shaped by the methodology that is selected. This can cause problems for multi-method IAs which may not be able to apply a single definition for a marker for each of the methods used. In addition, impact assessors should always seek to keep the number of variables they measure to a manageable number and not be tempted to go for a comprehensive approach that will impact adversely on data quality and study relevance.

4. ASSESSING IMPACTS: THE METHODOLOGICAL MENU[5]

The commonest methods used in impact assessment are sample surveys, rapid appraisal, participant-observation, case studies and participatory learning and action (Table 2). A comparison of their strengths and weaknesses is presented in Table 3. This paper can only explore these comparisons in a limited way as a detailed analysis would fill a book.

---

[4] Gaile and Foster (1996:Annex 1) provide a vast list of all the variables measured in 11 recent studies of microfinance impacts. The reader will also find this very useful.

[5] This section draws heavily on the work of Montgomery et al (1996).

Over the last decade impact assessment studies have increasingly moved away from single method approaches (e.g.;Hossain 1988; Fuglesang and Chandler 1986) to mixed or pluralist approaches (e.g.;Hulme and Mosley 1996; Mustafa et al 1996). Recently the methodological menu has been extended with the introduction of participatory learning and action (PLA) and participatory impact assessment (PIA). The choice facing microfinance IAs today is not 'what method should we choose', but 'what mix of methods should we choose and how should we combine them'?

Table 2: Common Impact Assessment Methods

Method	Key Features
Sample Surveys	Collect quantifiable data through questionnaires. Usually a random sample and a matched control group are used to measure predetermined indicators before and after intervention
Rapid Appraisal	A range of tools and techniques developed originally as rapid rural appraisal (RRA). It involves the use of focus groups, semi-structured interview with key informants, case studies, participant observation and secondary sources
Participation Observation	Extended residence in a program community by field researchers using qualitative techniques and mini-scale sample surveys
Case Studies	Detailed studies of a specific unit (a group, locality, organisation) involving open-ended questioning and the preparation of 'histories'
Participatory Learning and Action	The preparation by the intended beneficiaries of a program of timelines, impact flow charts, village and resource maps, well-being and wealth ranking, seasonal diagrams, problem ranking and institutional assessments through group processes assisted by a facilitator

Table 3: Comparative Strengths and Weaknesses of Different Methods

Method Criteria	Surveys	Rapid appraisal	Participant observation	Case studies	Participatory Learning and
1. Coverage (scale of applicability	High	Medium	Low	Low	Medium
2. Representativeness	High	Medium	Low	Low	Medium
3. Ease of data standardisation, aggregation and synthesis (eg. quantification)	High	Medium	Medium or Low	Low	Medium or low
4. Ability to isolate and measure non-project causes of change	High	Low	Low	Low	Low
5. Ability to cope with the attribution problem	High	Medium	Medium	Medium	Medium
6. Ability to capture qualitative information	Low	High	High	High	High
7. Ability to capture causal processes	Low	High	High	Medium	High
8. Ability to understand complex processes (eg. institution building)	Minimal	Medium	High	Medium	Medium
9. Ability to capture diversity of perceptions	Low	High	Very high	Medium	High
10. Ability to elicit views of women and disadvantaged groups	Low	Medium	High	High (if targeted)	Medium

Method Criteria	Surveys	Rapid appraisal	Participant observation	Case studies	Participatory Learning and
11. Ability to capture unexpected or negative impacts	Low	High	Very high	High	High
12. Ability to identify and articulate felt needs	Low	High	High	Medium (due to low coverage)	High
13. Degree of participation encouraged by method	Low	High	Medium	Medium	Very high
14. Potential to contribute to stakeholder capacity building	Low	High	Low	Medium to low	Very high
15. Probability of enhancing downwards accountability	Low	High	Medium	Medium	High
16. Human resource requirements	Specialist supervision, large numbers of less qualified field workers	High skilled practitioners, who are able to write-up and analyse results	Medium skilled practitioners, with good supervision, who are prepared to commit for lengthy period	Medium skilled practitioners with good supervision	High skilled practitioners
17. Cost range	Very high to medium	High to medium	Medium to low	Medium to low	High to Medium
18. Timescale	Very high to medium	Medium to low	High	High to medium	Medium to low
	• Surveys	• Rapid appraisal	• Participant observation	• Case studies	• Participatory learning and Action

Source: Adapted from Montgomery et al (1996).

The starting point to answer this question is to determine the degree of scientific validity an IA is to achieve. A scientifically valid design is required when

- # it is essential to 'prove' impact for policy and investment purposes;[6]
- # cost effectiveness is a major issue;
- # the costs of a program are very high;
- # the program 'model' is controversial (ie; the claimed results are much higher than informed observers would expect); and
- # other donors, governments and organisations are to be influenced.

If such a 'robust' design (Montgomery et al 1996) is required that can deal with the problems of attribution and fungibility in detail then a sample survey, preferably cross-checked by other methods, must be mounted. This will be costly and time-consuming, but if well conducted will permit a defined degree of confidence in the results to be achieved. (Table 3). By contrast, if an IA is required to provide corroboration of program impact and help in strengthening aspects of implementation, then a mix of rapid appraisal and a small-scale survey are likely to be more appropriate. This will be relatively cheap and can yield results within a short time frame. These methods must be used with rigor, however, and require systematic design and the achievement of relatively high standards. Table 4 provides a summary of conditions under which different methods are and are not appropriate (for a fuller discussion see Montgomery et al 1996). The IA designer must be aware that s/he is not choosing an optimal method but trying to achieve a good 'fit' between IA objectives, budget and timescale with methods.

[6] As the CGAP virtual meeting observed, genuinely 'proving' anything about MFI impacts is infeasible. The key issue is the level of plausibility of conclusions. In the situation specified above this would be reflected in statistical testing reaching a 95 per cent, or more, degree of confidence.

Table 4: Conditions in which Key Methods are, and are not, Appropriate

Sample Surveys are appropriate when:	Rapid Appraisal and/or PLA are appropriate when:	Participant Observation and/or Case Studies are appropriate when:
A project affects large numbers of beneficiaries Policymakers require accurate estimates of project impacts Statistical comparisons must be made between groups over time and/or between locations Project delivery/implementation mechanisms are operating well, thereby justifying investment in the assessment of impacts The target population is heterogeneous and it is difficult to isolate the influence of factors unrelated to the project (e.g. contextual variables, other programmes etc)	The project is adopting or promoting participatory principles in (re-)planning, implementation, monitoring and evaluation An understanding of the motivations and perceptions of project clientele is a priority One of the purposes of the study is to assess whether or not felt needs are being addressed by the project The impact of community-based organisations or other institution building activities are of importance There is a need to understand the quality of other data collected through surveys There is a need for contextual studies before designing more complex monitoring or impact assessment exercises (eg. case studies, or surveys).	An understanding of the motivations and perceptions of project clientele is a priority Other methods (surveys and rapid appraisals) are unlikely to capture the views of minorities or women One of the purposes of the study is to assess whether or not felt needs are being addressed by the project The impact of community-based organisations or other institution building activities are of importance There is a need to understand the quality of other data collected through surveys or rapid appraisals (eg. causal processes) There is a need for contextual studies before designing more complex monitoring or impact assessment exercises (eg. before carrying out rapid appraisals, or before designing a survey)
Sample Surveys are usually not appropriate when:	**Rapid Appraisal and/or PLA are not appropriate when:**	**Participant Observation and/or Case Studies are usually not appropriate when:**

A project affects small numbers of beneficiaries Policymakers are mainly concerned with project outcomes. (Was infrastructure completed on time and within budget? How many people us the health clinics?) Project implementation is recent and untested, and it is likely that the way in which the project is implemented will have little impact at the present time. The purpose of the assessment is to study and evaluate complex activities or processes (eg. the development & operation of community-based organisations; qualitative use of skills as a result of training programmes) The purpose of the assessment is to document easily observable changes in the physical environment or other tangibles (which can be assessed through simpler, structured visits) The purpose of the assessment is to understand whether or not the project is meeting the felt needs of the project clientele	Projects are relatively un-complex, in which bounded locations are not units of analysis (eg. Health centres serving a wide catchment area, or large schools serving a variety of communities) Indicators of project impact are uncontroversial, and negative impacts are unlikely Standardised and statistically representative generalisations for a large and diverse project population are regarded as the sole priority Participation of clientele is not a priority (eg. in public administration or power sector reform, or an organisational change programme - in these types of projects more limited focus group discussions with staff may be more appropriate)	The project is small and "uncomplicated", providing a specific service or limited intervention which is unlikely to affect community dynamics beyond a few specific effects (eg. disease specific health facilities or campaigns) Bounded locations are not units of analysis (eg. health centres serving a wide catchment area, or large schools serving a variety of communities) Indicators of project impact are clear and easily measurable or assessable (by survey or rapid appraisals) Indicators of project impact are uncontroversial, and negative impacts are unlikely Information is needed quickly, and standardised, representative (statistical) generalisations are regarded as the sole priority

Source: Adapted from Montgomery et al 1996.

5. CAUSES AND EFFECTS: THE PROBLEMS OF ATTRIBUTION AND FUNGIBILITY

The major conceptual problems that confront the IA of microfinance relate to attribution and fungibility. At the heart of impact assessment is the attribution of specific effects (ie impacts) to specific causes (ie interventions). From the vast literature on microfinance IA it is possible to draw out three very different ways by which authors seek to demonstrate attribution. The first is the conventional scientific method with its origins in the natural sciences. The other has its roots in the humanities (particularly history and anthropology) and focuses on making a reasoned argument supported by theory and specific pieces of evidence. Although the former has tended to dominate discussions about microfinance impact assessment (see for example the studies reviewed by Gaile and Foster 1996) the latter tradition is being increasingly used (Bouman and Hospes 1994; Ardener and Burman 1995; Remenyi 1991). The third part of this section explores a recent entrant to the field - participatory learning and action - which offers a radical challenge to both conventional IA and to 'science' itself. Although these three approaches can be separated for analytical purposes, in practice most studies weave elements of these approaches together.

5.1 Scientific Method

Scientific method seeks to ensure that effects can be attributed to causes through experimentation.[7] A particular stimulus to a particular object in a rigorously controlled environment is judged to be the cause of the observed effect. The experimental approach is virtually infeasible in the social sciences, because of the nature of the subject matter, and so the approach has been adapted into quasi-experiments (Casley and Lury 1982). Quasi-experiments seek to compare the outcomes of an intervention with a simulation of what the outcomes would have been, had their been no intervention. One method for this is multiple regression, but this has rarely been used in microfinance IA because of its enormous demands for data on other possible causal factors and its assumptions (Mosley 1997:2-3). A second approach is the *control group* method which has been widely used. This requires a before and after comparison of a population that received a specific treatment[8] (ie a microfinance program) and an identical population (or as near as possible) that did not receive the treatment. While this idea is elegantly simple a number of 'elephant-traps' may befall its user. In particular problems of sample selection bias, mis-specification of underlying causal relationships and respondent motivation (see later) must be overcome (also see Appendix 6 for a summary of problems and possible corrective measures).

Selection bias may occur because of

a. difficulties in finding a location at which the control group's economic, physical and social environment matches that of the treatment group;

[7] This section draws heavily on Mosley's (1997) excellent background paper, which the reader should consult.

[8] This method is used widely in the medical and agricultural sciences.

b. the treatment group systematically possessing an 'invisible' attribute which the control group lacks (most commonly identified as entrepreneurial drive and ability);

c. receiving any form of intervention may result in a short-term positive response from the treatment group (the Hawthorne effect);

d. the control group becoming contaminated by contact with the treatment group (though this could be a long term program goal!); and

e. the fungibility of the treatment (e.g.;when a loan is transferred from a borrower to someone else or when the loan is not used in the planned way).

Problems (a) and (d) can be tackled by more careful selection of the control group. This applies particularly to controlling for access to infrastructure (which has a key influence on input and output prices as well as other variables) and ensuring that the control group is located far away from the treatment group. Problems (b) and (c) are more intractable, but in many cases they can be tackled by using accepted 'clients-to-be' who have not yet received microfinance services as the control group (Hulme and Mosley 1996, chapter 4). It must be noted, however, that this approach will not be valid when the take up of microfinance services is based on diffusion through a heterogeneous population.[9]

This leaves the problem of *loan fungibility*. This can be seen as an intractable problem as '...no study has successfully controlled for the fungibility of resources between the household and the assisted enterprise' (Gaile and Foster 1996:24). Using case study materials to cross-check actual loan use against intended loan use and thus estimating 'leakage' is one possible approach to controlling for fungibility (Pulley 1989; Mosley 1997). However, for all studies except those that focus exclusively on 'the enterprise', then a concern about fungibility may be irrelevant. For studies looking at the household, the community or the household economic portfolio (see Section 3.2) fungibility is not a problem for the assessor, rather it is a vital strategy for the client. The best investment returns may be on 'consumption' (in terms of developing or maintaining human capital through school fees and doctors' bills, or buying food at a time of crisis when the credit terms on 'in-kind' borrowing from traders may be exceptionally high). From this perspective the task of the assessor is not to pretend that microenterprises are 'firms' whose inputs and outputs can be precisely identified and measured but to recognise that the impacts of microfinance must be assessed at a variety of levels. The assessor attempting to control for fungibility (to prove impact) has failed to recognise that fungibility is a process to be encouraged (to improve impact)!

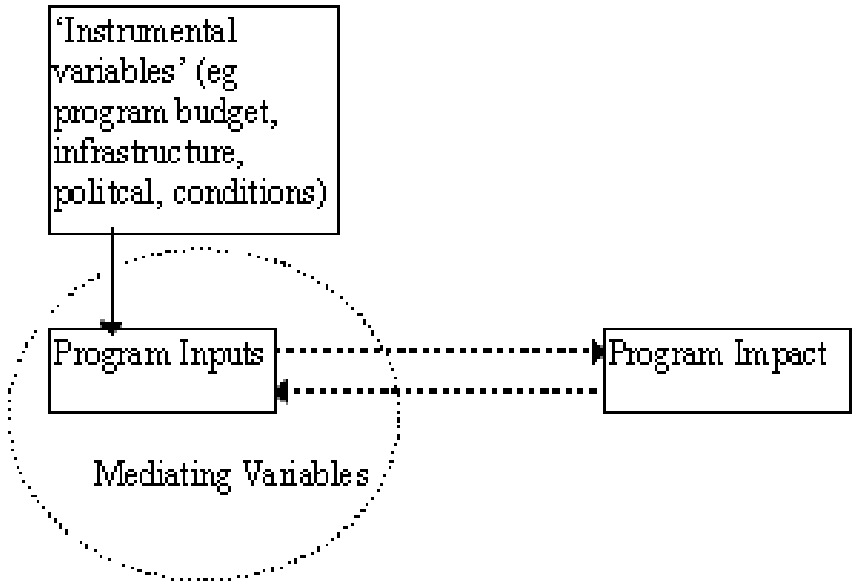
The mis-specification of underlying causal relationships arises most commonly because of the assumption that causality is a one-way process (Figure 1). This is a reasonable assumption with regard to natural materials (though it does not go unchallenged by contemporary philosophers of science). For

[9] We were not able to use this method for the Thrift and Credit Cooperatives in Sri Lanka, as the 'early' members of a cooperative are generally drawn from higher income/assets/status groups while 'late' joiners are from lower income/assets/status groups. In this case the use of a 'clients-to-be' control group would have led to an exaggerated assessment of the economic impacts.

human activity it is commonly invalid, as causation may also run from impact back to intervention. Mosley (1997:6) illustrates this with the example of a program whose field staff put pressure on a borrower to repay her loan; this may succeed in the short-term but may induce the borrower to sell assets (machinery, land, trees) which reduce the probability of repayment in the longer-term. Such reverse causation need not necessarily be negative and, from the perspective of more process-oriented analytical frameworks, is essential if programs are to continually learn from their experience and improve (rather than prove) their impact.

Such problems can be overcome by the adoption of models that conceptualise causation as a two-way process (Figure 3), the use of two-stage least squares technique and regression analysis (ibid:7). Such an approach is enormously demanding in terms of data requirements, technical expertise and costs. It will only be feasible on very rare occasions (for example see Khandker 1996).[10] For most researchers adopting the scientific method, reverse causality is a problem to be coped with rather than overcome. The main means of dealing with it are (i) tracing dropouts from both the treated and control groups; (ii) only conducting IAs on relatively mature programs; (iii) interim impact monitoring activities to gather qualitative information about the complexity of causality; and (iv) retrospective in-depth interviews with clients (ibid:6).

Figure 3: Simultaneous Causation Between Intervention and Impact



Source: Mosley 1997:7

[10] The scale of this survey was so great that in Bangladesh it was known as 'the mother of all surveys'. It operated on such a scale that other researchers found difficulties in conducting fieldwork during the period of operation as research institutes, local researchers and enumerators and data analysis facilities were all 'booked up'.

5.2 The Humanities Tradition[11]

The broad set of approaches that fall under this head have their roots in the humanities. Originally history and geography were the 'lead' subjects, but over the last 20 years anthropology has become the hearth. Its main features are an inductive approach, a focus on key informants, recording by notes or image and the data analyst is usually directly and heavily involved in data collection. This tradition does not try to 'prove' impact within statistically definable limits of probability. Rather, it seeks to provide an interpretation of the processes involved in intervention and of the impacts that have a high level of plausibility. It recognises that there are usually different, and often conflicting, accounts of what has happened and what has been achieved by a program. The validity of specific IAs adopting this approach has to be judged by the reader on the basis of (i) the logical consistency of the arguments and materials presented; (ii) the strength and quality of the evidence provided; (iii) the degree of triangulation used to cross-check evidence; (iv) the quality of the methodology; and (v) the reputation of the researcher(s). Whether 'standards' could be specified for such work - to help its users appreciate how rigorously designed they are - is an important issue raised in the conclusion. Commonly the bulk of data generated by such an approach is 'qualitative', although at later stages of analysis such work often quantifies some data. The main types of methods used have been discussed in Section 4 (in particular see Table 2). Chao-Beroff (1997) also identifies the use of 'life stories' in IA studies.

Although such work has been common in development studies for decades, it is only recently that its relevance for IA has been recognised.[12] This recognition has arisen partly because of the potential contribution of qualitative approaches (especially in understanding changes in social relations, the nature of program staff-beneficiary relations and fungibility) and partly because of the widespread recognition that much IA survey work was based on inaccurate information collected by questionnaire from biased samples (Chambers 1993). Low budget and low rigour IAs adopting the scientific method were at best pseudo-science, but more often simply bad science, despite the sophisticated analytical tools that were utilised!

However, IAs with their roots in the humanities still have considerable difficulties with regard to the attribution of cause and effect. Such studies cannot usually demonstrate the causal link as they are not able to generate a 'without program' control group although at times the work of some researchers neglects to mention this to the reader. Instead, causality is inferred from the information about the causal chain collected from intended beneficiaries and key informants, and by comparisons with data from secondary sources about changes in out-of-program areas. Problems also arise because not infrequently the labels

---

[11] Commonly such work is referred to as 'qualitative', but the quantitative/qualitative dichotomy is a false dichotomy. Most quantitative studies extract qualitative data from the respondent, have the interviewer immediately codify or quantify it and then use only numerical analysis. Many 'qualitative' studies transform their data into quantities at later stages of analysis (Moris and Copestake 1993:4).

[12] As Moris and Copestake (1993:1) point out 'the much recommended text on data collection by Casley and Lury (1982)... included two cursory paragraphs on 'quick and dirty' techniques... almost half of the World Bank's publication that superseded it (Casley and Kumar 1988) is concerned with qualitative methods'. Studies using this approach include Hulme and Mosley (1996), Rutherford (1993), Fuglesang and Chandler (1986 and 1993). For details of many recent examples see Mayoux (1997).

'rapid appraisal', 'mini-survey' and 'case study' are applied to work which has been done in an *ad hoc* manner and does not achieve a minimum professional standard in terms of the rigor of data collection and analysis. Examples of this include: i) basing data collection only in program areas that are performing well, and surveying best clients, and ii) inferring that the data collected applies to all clients.

While such studies cannot provide the degree of confidence in their conclusions that a fully resourced scientific method approach can yield, my own judgement is that their conclusions are often more valid than much survey based IA work that masquerades as science but has not collected data with rigor. Whatever, it is becoming increasingly common to combine 'scientific' and 'humanities' approaches so as to check the validity of information and provide added confidence in the findings (for example Hulme and Mosley 1996; Schuler and Hashemi 1994; Hashemi et al 1995). In the future dealing with attribution by multi-method approaches seems the way forward.

5.3 Participatory Learning and Action (PLA)

In the last five years participatory approaches to development planning and management have moved from being a fringe activity to centre stage. While many donor agencies, have simply added a bit of PLA to their existing procedures, it can be argued that this is inappropriate as conceptually participatory approaches challenge the validity and utility of the scientific method as applied to developmental problems (Chambers 1997). According to this line of argument the scientific method fails as: it ignores the complexity, diversity and contingency of winning a livelihood; it reduces causality to simple unidirectional chains, rather than complex webs; it measures the irrelevant or pretends to measure the immeasurable; and, it empowers professionals, policymakers and elites, thus reinforcing the *status quo* and directly retarding the achievement of program goals. At heart, PLA theorists do not agree that ultimately there is one objective reality that must be understood. Rather, there are multiple realities and before any analysis or action is taken the individuals concerned must ask themselves, 'whose reality counts?' (ibid). The answer must be that the perceived reality of the poor must take pride of place as, if development is about 'empowering the poor' or 'empowering women' (as virtually all development agencies now say), then the first step towards empowerment is ensuring that 'the poor' or 'women' take the lead in knowledge creation and problem analysis.[13]

For impact assessment the purist PLA line is damning - '...conventional baseline surveys are virtually useless for impact assessments... The question now is how widely local people can be enabled to identify their own indicators, establish their own participatory baselines, monitor change, and evaluate causality...' (ibid:123). By this means two objectives may be achieved (i) better impact assessments, and (ii) intended beneficiaries will be '...empower[ed] through the research process itself' (Mayoux 1997:2). In practice, the art of participatory impact assessment (PIA) is in its infancy and a pragmatic rather than a purist approach has been common (see Chao-Beroff (1997) for an example of an NGO's use of participatory methods and Martyn-Johns (1996) for a comprehensive review of PIA).

[13] The reader wishing to explore PLA and PRA (participatory rural appraisal) is referred to Chambers (1997) as there is not space to more fully explore these ideas in this paper. See Mayoux (1997) for a discussion of empowerment as a program goal, with particular reference to gender.

The reliability of participatory methods varies enormously, as with 'scientific' surveys, depending '...largely on the motivation and skills of facilitators and those investigated and the ways in which informants' perceptions of the consequences of research are addressed' (Mayoux 1997:12-13). Nevertheless, it is argued that '...a number of rigorous comparative studies have shown that, when well-conducted, participatory methods can be more reliable than conventional surveys' (ibid and see Chambers 1997:141-146).

To date the literature on PIA has only partially addressed the issue of attribution. From a scientific perspective PIA has grave problems because of the subjectivity of its conceptualisations of impact; the subjectivity of the data used to assess impact; because the variables and measures used vary from case to case and do not permit comparison; and, because its pluralist approach may lead to a number of mutually conflicting accounts being generated about causality. From the perspective of a 'new professional' (Chambers 1997) then such a set of accounts is unproblematic, as it reflects the complexity and contingency of causality in the real world. In addition, PIA, it can be argued, contributes to program goals (perhaps particularly in terms of empowering women (Mayoux 1997) and the poor) by not facilitating the continued dominance of target groups by powerful outsiders. Why dwell on issues of attribution when efforts to overcome such problems require the adoption of methods that will undermine the attainment of program goals?

6. PRACTICAL PROBLEMS OF IMPACT ASSESSMENT (AND THEIR SOLUTIONS)

While IA design is intellectually challenging, effective impact assessment is also challenging in terms of implementation. Here I cover some of the main problems.[14]

6.1 Costs and Confidence

The design of an IA must be very closely related to the budget available: this may be a platitude but overambitious designs continue to lead to poor IA studies. Interestingly, in this age of cost consciousness, the literature on microfinance provides no specific information on the overall or unit costs of IA studies of microfinance; 'high', 'medium' and 'low' are about as good as the data gets![15]

From verbal reports it is clear that IAs adopting the scientific method and seeking to 'prove' impact cost the earth (say US$1 million to US$5 million depending on the number of MFIs studied). At the other extreme high quality, rapid appraisals of the impact of individual schemes by gifted and knowledgeable

[14] For problems such as respondent recall capacity, the valuation of subsistence products and seasonality, the reader should consult standard texts (e.g.;Moser and Calton 1972; Casley and Kumar 1988; Devereux and Hoddinott 1993).

[15] The only report that explicitly identifies absolute costs and relative costs (as a percentage of program budgets) is Montgomery et al (1996) mainly in relation to natural resource and social programs. It reports that the 1994 IA of BRAC's credit program cost US$250,000.

individuals can produce useful findings on 'improvement' for relatively small sums (around US$2000 to US$6000). Between these two extremes are a vast array of different options. A rapid reading of the literature produces the following points.

(a) Studies intended to produce authoritative evidence of impact using the scientific method will be rare exceptions in the IA field. Their costs mean that objectives must be very carefully determined prior to design.

(b) The idea that 'qualitative' and 'participatory' assessments methodologies are cheap needs to be challenged (Mayoux 1997). While such approaches are much cheaper than vast surveys, rigorous qualitative IAs will require the use of high quality staff who are given time to prepare properly. Costs of tens of thousands of dollars, rather than thousands, should be anticipated.

(c) For studies of moderate budget (ie most studies) the best approach to ensuring the validity of findings will be through triangulation and using a mix of survey, qualitative and participatory techniques. The alternative, of trying to achieve a representative sample size on a limited budget, is likely to lead to severe losses in the quality of data and/or the representativeness of the sample.

(d) Limited investments in project monitoring by program staff make moderate cost impact assessment at high levels of quality much more feasible as less primary data collection is necessary (see Montgomery 1996 and later parts of this paper).

6.2 Respondent Motivation

A 'rational actor' confronted by an impact assessor asking standard IA questions ('what is your income? what do you spend your money on? how do you get on with your husband?') would soon tell the interviewer where to put his/her survey instrument. Fortunately, in the world of practice, more polite responses are the norm but the issue of how to persuade respondents to spare the time for an interview, and provide accurate and honest answers, is an important one that is rarely mentioned in IA methodological statements. Different strategies are needed for different types of respondent - program beneficiary, control group and program drop out. As a rule of thumb many researchers suggest that one and a half hours should be seen as the maximum possible length for an interview.

Beneficiaries are the easiest group to approach as generally they accept 'answering questions' as part of the package of being in a program or dealing with an MFI. Motivation can be enhanced by having interviewers introduced by program officers: but, this has the danger of linking the assessor with field level staff and encouraging the recounting of 'the right answers'.[16] For both data quality and ethical reasons the

[16] At one of our large Asian case study institutions (Hulme and Mosley 1996, Vol. 2), program field staff visited villages that we had randomly selected for survey and told borrowers to make sure they gave our interviewers 'the right answers'. Fortunately, qualitative research revealed this and other villages were selected for

personal introductions that interviewers make prior to interview need to be carefully worked out so that respondents understand why they are being questioned.

Motivation is a more difficult issue with control groups as, having by definition no connection with a program, they have no incentive to cooperate. In many cases, however, the novelty and amusement value of being interviewed is sufficient encouragement (though expatriates should note that the willingness of people to be interviewed may be higher than is the norm when they are working at a field site because of their rarity value). The problems of response increase significantly if longitudinal data is collected as second and third interviews have much less amusement value. In such cases rewarding interviewees should be considered to promote data quality and for ethical reasons (what right have impact assessors to assume that the opportunity costs of an interview, particularly for poor people, are zero?). This can take the form of a social reward, such as bringing soda waters and snacks to share with respondents (I have found this works well in East Africa), or 'bribery' (Mosley 1997:8) where the interviewee is paid cash for surrendering her/his time.[17]

Program drop-outs represent a particular problem, and a failure to pursue drop-outs may have led to our research underestimating the negative impacts of microfinance (Hulme and Mosley 1996). When the drop-out is traceable then significant effort is merited to obtain an interview/re-interview. Where drop-outs cannot be traced, or death has occurred, then a replacement respondent sampled at random from the original population, and preferably from the same stratum, should be interviewed (see Mosley 1997:7-8).

Participatory and rapid appraisal methods that work with groups generally manage to muster respondents because of the social interaction they create. However, care needs to be taken to observe who has turned up and, perhaps more significantly, who has not come to the meeting (Mayoux 1997). Additional interviews or focus groups may be necessary to collect information from people who do not turn up for communal PLA or RRA sessions.

6.3 Consultant and Interviewer Quality and Availability

In many countries recruiting IA personnel who have the skills and qualities to interview, collate, analyse and write up work is a key problem that applies at both consultant and fieldworker levels. A number of ways of trying to deal with this problem, or relieve it in the longer term can be identified.

(a) If IAs are scheduled well in advance then the likelihood of 'booking up' quality personnel is enhanced. The donor preference for last minute contracting (just out of time management) leads to the recruitment of weak IA teams and subsequent problems with IA logistics and quality.

---

survey!

[17] This method is widespread in the USA and UK when market researchers convene focus groups to test new products. It should be noted that once interviewees in an area are paid for interviews then the likelihood of non-cooperation in the future, unless fees are paid, is greatly increased.

(b) Longer term, donors keen on IA need to think how they can strengthen the human and institutional resources available for IA in developing countries. This could involve technical assistance to research institutes and/or private consultancy groups; the management of a flow of work so that people can see IA in career terms (rather than a few dollars income on the side); moves to produce IA standards and/or codes of practice that will help to professionalise impact assessment; and simple manuals and 'case study methodologies' that will help assessors to achieve best practice more quickly.

(c) The information collected by survey methods can generally be transferred between analysts without too much difficulty. By contrast, the data collected by qualitative methods usually entails that the principal researcher does the write-up and so contracts for qualitative researchers should ensure their availability for writing-up.

6.4 The Problem of Low Impact-Impact Assessments

A final problem of IA concerns the impact of IAs. This depends in part on the original objectives of a study. It applies to both 'proving' and 'improving' IAs. The evaluation literature of the 1980s bemoans the limited influence of evaluation on subsequent decision-making. IA has inherited this problem, as illustrated by the very limited influence of large scale impact assessment studies (Mustafa et al 1996) on the microfinance activities of the Bangladesh Rural Advancement Committee (BRAC).

A number of ways of ameliorating this problem can be identified.

(a) Impact assessors need to devote more time to the 'use' of their studies (and perhaps a little less time to the product itself!). Their focus must go beyond 'the report' into a dissemination strategy aimed at decision-makers: bullet point summaries, snappy presentations and strategic cups of coffee are the key to this environment.[18]

(b) The timing of findings needs to be carefully considered. As a general rule of thumb the longer the length of time between data collection and findings presentation, then the lower the impact for IAs focused on 'improving' practice. The common response to initial findings presented more than nine months after completion of fieldwork is 'our program has already been redesigned so your findings have little relevance'.

(c) Program managers often regard impact assessors as impractical people who have lots of time on their hands. For high cost approaches pursuing the scientific method this will be of only limited significance as the people to whom one's results must be credible are in Washington and European capitals. However, for the vast majority of IA studies the issue of how to develop constructive relationships with program staff requires careful thought and action. Efforts to achieve co-ownership of findings by involving program staff in IA

[18] The Khandker (1996) study and its related papers is an obvious example. Microcredit specialists with PhDs in Economics describe the study as 'virtually unreadable' and thus of little use to policy makers.

design, showing respect for their ideas and opinions, and discussing interim findings are possible ways of making influence more probable.

7. SELECTING THE APPROACH: ACHIEVING 'FIT'

The key task for the IA designer is to select an approach that can meet the objectives of the specific assessment at an acceptable level of rigor, that is compatible with the program's context, that is feasible in terms of costs, timing and human resource availability and that avoids the problems identified in earlier sections. Wherever possible an IA methodology should be piloted before full implementation. The questions that s/he must answer can be summarised as follows.

(1) What are the objectives of the assessment?

(2) How is the information to be used and by whom?

(3) What level of reliability is required?

(4) How complex is the program, what type of program is it, what is already known about it?

(5) What resources (money, human and time) are available?

The range of specific responses to these questions is infinite, but for the purposes of this paper they are grouped into four categories. These categories are based on Little (1997) but their characteristics have been substantially modified. These range from impact monitoring and validation, through simple and moderate approaches to more complex approaches (Table 5). These can be viewed as a hierarchy, but there is a great danger in this as this may infer that complex approaches are best.

Table 5: Different Approaches to IA and their Key Characteristics

	Key Characteristics							
Methodological Approach	**Objectives**	**Main Audience**	**Level of Reliability**	**Main Method**	**Other Methods**	**Costs**	**Timescale (start up to dissemina-tion**	**Other**
1 Impact Monitoring (don't do an IA)	To improve programme performance	Senior programme managers and field managers	Moderate to low	Use of available data through a mix of methods	N/A	Low to Moderate	2 to 6 months initially, but then on a continuous basis	Informal impact monitoring systems already exist in most MFIs. Strengthening these is the lowest cost route to relating impacts to action. If formalised, these systems make IASs much easier
2 Simple	To test existing under-standing of impacts and improve programme performance	Senior programme managers and donor reps, in-country	Moderate to low	No main method -	Many methods used	Low to moderate	9 to 18 months	If empowerment and local institutional development are key programme goals then a participatory approach may be needed
3 Moderate	Proving impact to a reasonable degree of reliability and improving programmes	Bilateral and recipient policy-makers and managers and senior operational managers	Moderate*	Sample survey with control group	Rapid appraisal Case studies P-O	High	15 to 30 months	If empowerment and local institutional development are key programme goals then a participatory approach may be needed
4 Complex	Proving programme impact	Econometricians ,economically literate policy analysts in donor and recipient countries and academic researchers	High*	Large sample survey with control group	Background data for contextual-isation	Very high	48 months plus	Of little relevance for programme improvement but important for investment programming and donor accountability. Ethical questions about whether OECD nations apply such high standards to themselves.

* Reliability ultimately depends on the quality of fieldwork: this can lead to poorer levels of reliability than econometric analysis might claim

7.1 Impact Monitoring and Validation (or, do not do an IA!)

Commonly the answer to the above questions should be 'don't proceed with an impact assessment' (as a program's emphasis on 'institution building' will be undermined by IA and/or the resources are not available). Unfortunately donor agencies generally lack the courage to reach this decision and consultants (*mea culpa*) may have a vested interest in not promoting this option. Instead, donors could focus on strengthening the internal impact monitoring capacities of the microfinance institution and occasionally checking the quality of this information by using external monitors to validate it.

Building internal impact monitoring capacities does not mean creating a large impact assessment unit within an MFI. Rather, it means helping the MFI develop its MIS and the work of its internal monitoring and research units to collect readily available data (outreach, repayments, drop-out rates etc) alongside 'simple to gather' forms of data collected by the internal monitoring unit on who is using services, what for, why and what they like/dislike about the services. Much of this work can be done by focus groups, short interviews and rapid appraisal.[19] It may also entail assistance to MFIs in how to use short-term consultants to best effect.

These systems already operate in some of the large Asian MFIs and are the basis upon which their directors take many 'improvement' decisions. Strengthening these systems and occasionally verifying them - rather than financing more complex external impact assessments - is probably the best way to achieve the 'improving' goals of IA. The types of verefication process used in Social Audits (New Economics Foundation 1996; Zadek and Gatward 1996) provide a model of ensuring that internal impact assessments are valid.

7.2 A Simple Approach

This would seek to provide timely information at a relatively low cost about program impacts. These are the most numerous forms of IAs. Reliability is moderate, at best (and based mainly on triangulation), and the major objective is to test the existing understanding of impacts and contribute to improvements in program operation. The main audiences are program managers and donor 'country-based' staff. The central methodological feature of such an approach is the use of a variety of methods. Usually this involves a small scale client survey, compared with a comparison group that could be rapidly identified (e.g.;approved clients who have not yet received services), and cross-checked by rapid or participatory appraisal methods. If a baseline study is not available then a recall methodology would be utilised. The key variables to be studied would depend on program objectives, but for income and assets the focus would be on ordinal and nominal measurements (see Little 1997:17). For programs prioritising empowerment goals and local institutional development, then participatory methods would be highlighted and the survey work might be dropped altogether.

[19] See Montgomery et al (1996) for an excellent discussion of the role of impact monitoring vis a vis impact assessment.

Several participants in the CGAP vitual meeting (Gaile 1997) argued that improving the credibility, utility and cost effectiveness of simple approaches was where the greatest gains in IA could be realised. Characteristics to enhance the effectiveness of simple IAs include:

(i) focusing on a small set of key hypotheses;

(ii) using variables that have a 'track record' from earlier IA studies;

(iii) clearly documented use of traingulation;

(iv) methods applied consistently over time;

(v) use of small sized comparison group; and

(vi) careful training of IA staff (ibid:5).

7.3 A Moderate Approach

The moderate approach would involve substantially more costs than the simple approach, would yield higher levels of reliability (statistical inference rather than triangulation) and is not likely to deliver findings for a period of 2 to 3 years. Its focus is on both proving impact and improving programs. Its audiences would include policymakers (looking for reassurance about their agency's investments) and the senior managers of programs. The 'mix' would centre on a significant survey that would stratify clients and compare them with a carefully matched control group. The survey would involve at least two visits with a minimum of 12 months between them and recall techniques would not be used. Contextual and cross-checking materials would be produced by rapid appraisal techniques and carefully planned participant observation and case studies might also be commissioned. While the selection of variables would depend on program objectives the income and assets data would be extended (ibid:20) and measurement would focus on interval and nominal scales.

7.4 A Complex Approach

The complex approach focuses on ensuring high levels of reliability with regard to the attribution of causality and has an exclusively 'proving' orientation. Its main audiences are policymakers and researchers and it is likely to be 4 to 6 years after launch before findings are available. The central method in such an approach is a large scale sample survey very carefully constructed to represent all key features of the client population. This is compared against a carefully selected control group, so that the number of households surveyed is likely to be between 750 and 1500. At least 3 interviews will be conducted with each household over a period of 2 to 3 years. A much wider set of income and asset variables will be measured (ibid:23) and the focus will be on high precision through interval measurements. A set of related studies on institutional performance would be conducted, but the heart of the study would be the econometric analysis of survey findings. The budgets for such approaches are measured in millions of dollars.

8. CONCLUSION AND RECOMMENDATIONS

Drawing conclusions about how to improve the design of IA methodologies is, unfortunately, not simply about recommending an 'optimal' approach! Rather, it is about ensuring that an appropriate approach (Section 7) for the particular task in hand is chosen, and then, achieving the highest quality of study possible from that approach in the light of the context and resource constraints. To stimulate debate I have opted for a number of recommendations that derive from this paper and from the virtual meeting. The object is to address the messy question of how we can push present practice towards 'better practice': not to pretend that there is a clear and precise 'way forward'.

1. A strong case can be made that IA is too donor driven and externally-oriented. The possibility of using IA resources to strengthen internal MFI impact monitoring systems (and occasionally verify them) has not been sufficiently considered. The approach that many donors have to IA contradicts their declared goals of 'partnerships', 'stakeholders', 'institutional development' and 'empowerment'. (Some NGOs, such as SEWA and CDIR, are already attempting to integrate impact monitoring systems and impact assessment studies.)

2. The idea of methodological rigour applies to all IA designs and methods. Cost and complexity do not autamotically equate with rigour and, for example, a large survey can be poorly designed while a good PLA can yield highly accurate data about impact.

3. The main need in contemporary IA studies is for simple IA approaches that achieve a relatively high degree of plausibility. This can be done by using a mix of methods, focusing studies on a small set of key hypotheses, developing time series data, documenting triangulation, using the same methods for later studies and assessing comparison groups. Training IA staff and helping them to 'professionalize' (see recommendation 8) will be necessary to improve simple approaches.

4. Serious attempts to assess impact (whether simple, moderate or complex) must pursue the impact chain down to 'who' an MFI provides services to and 'what' the effects are on individual and/or household livelihoods. The Ohio State School approach (outreach plus institutional sustainability), or its extension into borrower transaction costs, provide useful insights but no grounds from which impacts on livelihoods can be inferred with plausibility. Similarly, studies that focus purely on impacts on the 'microenterprise' have limited value because of the permeability of the enterprise/household boundary.

5. The Household Economic Portfolio Model (HEPM), and variants on this approach, have much to recommend them when well resourced and complex approaches are being undertaken.

6. Problems of attribution can be greatly reduced by skilled survey design in complex IA approaches and by triangulation in other designs. Studies that focus on impact at several levels (e.g.;HEPM) can overcome many of the difficulties associated with fungibility.

7. A significant proportion of IAs have 'low impact' on policy and practice because assessors allocate too little time and energy to dissemination. Academic achievements depend on the book and econometric sophistry: practical achievements depend on providing clear and relatively simple messages to busy policymakers and managers.

8. Donor activity in the field of impact assessment has, to date, done little to strengthen the human resources and institutions (research institutes, consultancy groups and NGOs) in developing countries that conduct IAs. Institutional development in the IA field has strengthened OECD based institutions, with the notable exception of participatory methods. The long term benefits of work to improve IA methodologies are dependent on developing in-country design capacities.

9. The complexity of impact assessment means that improvements in IA must be based on raising the capacity of impact assessors, particularly the army of people who design and run simple IAs. This can be achieved by professionalizing and training assessors. Three key elements of this are: (i) developing a code of practice or set of standards for MFI assessors for different IA approaches; (ii) the production and dissemination of a manual for IA designers that provides a number of 'key checklists' and case studies of real world IA designs and methods (including questionnaires, control group selection, PLA facilitation and costs) and; (iii) establishing networks of MFI impact assessors/monitors in specific countries.

Appendix 1

Scope of Work for Discussion Paper on Impact Methodologies for CGAP

Terms of Reference

The objective of the discussion paper is to provide a document to be used as the basis for a discussion in a virtual meeting on impact assessment methodologies (April 7 - 12, 1997). The participants will include donors, technical experts and practitioners. They will be discussing the proposals in the paper, and the premises or assumptions on which these are based. The goal of the virtual meeting is to determine if there is agreement about a set of alternative methodologies for conducing impact assessments which address the credibility of the data vis a vis the intended use of the information and the objectives of the assessments.

The purpose of the discussion paper is to assess a range of methodologies which can be used to document the impact of microenterprise programs which take into account different a) program characteristics and context, b) objectives of an assessment, and c) intended primary use of the information. The discussion paper will draw upon the Background Papers as well as the knowledge and experience of the author. The consultant is to incorporate key points of each of the background papers in the discussion on methodological options.

The consultant will identify key categories related to methodology in order to incorporate the range of topics covered in the background papers. In covering different methodological issues, the discussion paper shall include the following topics:

- the use of survey instruments and/or qualitative approaches in undertaking formal/rigorous studies and less formal data collection exercises.

- the preference for longitudinal data. By and large the more rigorous impact assessments collect data more than once.

- the inclusion of a comparison group, thus permitting analyses between those who have accessed the microenterprise services and those who have not. The less rigorous studies which have focused only on program clients, do not permit one to analyse what changes might have occurred without use of the services. In the absence of control group information, little can be said about the relative advantages of program participation.

- accounting for the fungibility of resources within a household.

In addition, the following and other issues might be covered:

- recall periods vis-à-vis key variables and the credibility of the responses;

- recognition of the selection bias, both in terms of a program's clients and the location of the program. Location of a program can play a role in its success;

- temporal factors, including both the length of time that impacts are sustained, the point at which impacts first begin to occur, and the length of time in the program;

- the choice of analytical techniques which reflect the type of data collected; survey data permit analysis of the distribution of impacts whereas qualitative information permits analysis of dynamics, decisions and why changes occurred; and

- the identification of control variables such as gender.

In proposing methodological options for assessing the impact of microenterprise programs, the consultant should take into account in the feasibility of implementing these. Among the factors which should help to delineate the options should be cost, availability of resources (both within and outside the programs), and the utility of the proposed approaches vis a vis the objectives of the assessments and the utility of the findings to the implementing institutions. This information might be summarised in a table.

Until now impact assessments have tended to be seen apart from the implementation of programs, - off budget. However there is a growing awareness that impact assessments, especially baseline surveys, can be integrated into market surveys undertaken by implementing organisations to define and refine their product range. This should also be given consideration in the discussion of options. The consultant should explore whether acceptable standards for the methodological approaches to impact assessment might be set, vis-à-vis the objectives and intended use of the data.

Appendix 2

AIMS Documentation

Barnes, Carolyn (1996) *Assets and the Impact of Microenterprise Finance Programs,* Washington, DC:Management Systems International

Chen, M A and Dunn, E (1996) *Household Economic Portfolios*, Washington DC: Management Systems International

Dunn, Elizabeth (1996) *Households, Microenterprises and Debt*, Washington DC: Management Systems International

Dunn, E, Kalaitzandonakes, N and Valdivia, C (1996) *Risk and the Impacts of Microenterprise Services*, Washington DC: Management Systems International

Gaile, G L (1997) *Highlights and Recommendations from the Microenterprise Impact Virtual Meeting (Draft),* Washington DC: Management Systems International

Gaile, G L and Foster, J (1996) *Review of Methodological Approaches to the Study of the Impact of Microenterprise Credit Programs*, Washington DC: Management Systems International

Hyman, E and Dearden, K (1996) *A Review of Impact Information Systems of NGO Microenterprise Programs*, Washington DC: Management Systems International

Inserra, Anne (1996) *A Review of Approaches for Measurement of Microenterprise and Household Income*, Washington DC: Management Systems International

Little, P (1997) *Background Paper on Income and Assets as Impact Indicators*, Washington DC: Management Systems International

Sebstad, J and Chen, G (1996) *Overview of Studies on the Impact of Microenterprise Credit*, Washington DC: Management Systems International

Snodgrass, D (1996) *The Economic Policy and Regulatory Environment*, Washington DC: Management Systems International

For access to these use the website: http//:www.mip.org

Appendix 3

References

Ardener, S and Burman, S (1995) *Money-go-Rounds: the Importance of Rotating Savings and Credit Associations for Women*, Oxford: Berg

Bouman, F J A and Hospes, O (eds) (1994) *Financial Landscapes Reconstructed*, Boulder: Westview

Casely, D and Lury, D A (1982) *Monitoring and Evaluation of Agricultural and Rural Development Projects*, Baltimore: Johns Hopkins University Press

Chambers, R (1997) *Whose Reality Counts? Putting the First Last*, London: IT Publications

Devereux, S and Hoddinott, M (1993) *Field Research in Developing Countries*, London: Harvester Wheatsheaf

Fuglesang, A and Chandler, D (1986) *Participation as Process: What We Can Learn From the Grameen Bank*, Oslo: NORAD

Fuglesang, A and Chandler, D (1993) *Participation as Process - Process as Growth*, Dhaka: Grameen Trust

Gaile, G L (1997) *Highlights and Recommendations from the Microenterprise Impact Virtual Meeting (Draft)*, Washington DC: Management Systems International

Goetz, A M and Sen Gupta, R (1996) 'Who Takes the Credit? Gender, Power and Control Over Loan Use in Rural Credit Programs in Bangladesh', *World Development* 24(1), 45-64

Hashemi, S M, Schuler, S R and Riley, A P (1996) 'Rural Credit Programs and Women's Empowerment in Bangladesh', *World Development* 24(4), 635-654

Hulme, D and Mosley, P (1996) *Finance Against Poverty*, volumes 1 and 2, London: Routledge

Hossain, M (1988) *Credit for Alleviation of Rural Poverty: the Grameen Bank in Bangladesh*, Washington DC: IFPRI

Khandker, S R (1996) *Fighting Poverty with Microcredit: Experience of the Grameen Bank and Other Programmes in Bangladesh*, Washington DC: World Bank

Martyn-Johns, E (1996) *Participatory Impact Assessment: Putting Theory into Practice*, MA Dissertation, IDPM, University of Manchester

Montgomery, R, Davies, R, Saxena, N C and Ashley, S (1996) *Guidance Materials for Improved Project Monitoring and Impact Review Systems in India*, Swansea: Centre for Development Studies, University College Swansea

Moris, J and Copestake, J (1993), *Qualitative Enquiry for Rural Development: a Review*, London: IT Publications

Moser, C A and Kalton, G (1972) *Survey Methods in Social Investigation*, London: Heinemann

Mustafa, S et al (1996) *Beacon of Hope: an Impact Assessment Study of BRAC's Rural Development Programme*, Dhaka: BRAC

New Economics Foundation (1996) *A Social Audit Manual for Small Organizations*, London, New Economics Foundation

Peace, G and Hulme, D (1994) 'Microenterprise and Childrent:What are the Intra-household Impacts of Income-generating Programmes', *Small Enterprise Development* 5 (1), 21-29

Pulley, R (1989) 'Making the Poor Creditworthy: a Case Study of the Integrated Rural Development Program in India', *World Bank Discussion Paper* 58, Washington DC: World Bank

Remenyi, J (1991) *Where Credit is Due*, London: IT Publications

Rutherford, S (1993) 'Learning to lend: Informal Savings and Credit Schemes in Bangladesh' *SCF Working Paper No. 5*, London: Save the Children Fund.

Schuler, S R and Hashemi, S M (1996) 'Credit Programs, Women's Empowerment and Contraceptive Use in Rural Bangladesh', *Studies in Family Planning* 25(2), 65-76

Sebstad, J, Neill, C, Barnes, C and Chen, G (1995) *Assessing the Impacts of Microenterprise Interventions: A Framework for Analysis*, Washington DC: USAID

Yaron, J, Benjamin, M and Piprek, G (1997) *Rural Finance: Issues, Design and Best Practices*, Washington DC: Agriculture and Natural Resources Department, World Bank

Zadek, S and Gatward, M (1996) 'Transforming the Transnational NGOs: Social Auditing or Bust?', in M. Edwards and D. Hulme, *Beyond the Magic Bullet: NGO Performance and Accountability in the Post-War World*, West Hartford; Kumarian Press and London: Earthscan, 226-240

Appendix 4

Possible Corrective Measures for Improving the Accuracy of Quasi-experimental (Control-group) Evaluation Methodologies

Possible Corrective Actions	Supplementary Data Collection/design Modifications	Complementary Analytical Methods:	
Problems		**Informal but Cheap**	**Rigorous but Expensive**
1. Sample selection bias:			
(a) initial selection	Control for a larger number of determinants of welfare change (suggested list: initial income and assets, access to infrastructure, weather, land quality, gender, ethnicity) Randomisation of control location	RRA comparing asset changes etc. among borrower and non-borrower groups	Within-sample regressions using quantity of credit (or no, of loans) as independent variable
(b) recall methods	Cross-check components of income; check declared income against declared expenditure Also measurements of 'willingness to accept loan' in non-borrower sample, if the exercise described at right is wanted		Regress 'willingness to accept loan' on income within *non-borrower* sample, then use results of this regression to correct for 'greater willingness' to be 'entrepreneurial' among borrowers
(c) fungibility of capital (i.e. attribution)			Use other information (e.g. did borrower's stock of physical and human capital increase?

Possible Corrective Actions	Supplementary Data Collection/design Modifications	Complementary Analytical Methods:	
Problems		**Informal but Cheap**	**Rigorous but Expensive**
2. Inaccurate\ specification of causal mechanisms:			
(a) endogeneity	Instruments for propensity to accept a loan (e.g. capitalisation, gender)		2SLS regressions on borrower sample
(b) failure to pick up structural changes operating through time	Agricultural projects: collection of data over at least two cropping seasons Replacement of dropouts in both borrower and control groups (hence choose a larger sample than it is intended to initially interview	Case-studies within borrower group, in particular to assess technical interactions between borrower and control group	
3. Motivational problems:			
(a) persuading chosen control group to be interviewed	Cash or kind payments Use of borrowers-to-be as control group (*but beware: can only be used for periods before interviewee became a borrower)*		Multiple regression within borrower group
(b) Survey fatigue	Replacement of dropouts in borrower and control groups		
4. Resource costs	Keep borrower and control groups close together *(but risk of contamination of control through interaction with borrowers (see 2(b) above)*		

Source: Mosley (1997)